DOCUMENT RESUME

ED 056 539 FL 002 335

AUTHOR Mansell, Philip
TITLE Linguistic Parameters in Performance Models.
INSTITUTION Essex Univ., Colchester (England). Language Centre.
PUB DATE Dec 70
NOTE 31p.; In Occasional Papers 8, p51-81

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Articulation (Speech); Descriptive Linguistics; Environmental Influences; Language Patterns; Language Research; Language Universals; *Linguistic Competence; *Linguistic Performance; *Linguistic Theory; *Phonetics; *Phonology; Psycholinguistics; Speech; Syllables; Transformations (Language)

ABSTRACT

This paper deals with problems concerning the nature of the input to a phonetic processor. Several assumptions provide the basis for consideration of the problem. There is a phonological level of processing which reflects the sound structure of the language; the rules associated with it are not affected by variables associated either with the motor system or the environment. One level of processing can be called motor planning, another can be seen as executing the motor plan, and a level of control operates over the execution of the motor plan. The discussion that follows these assumptions concerns the phonetic processing mechanism and raises larger questions concerning the interrelation of linguistic competence and performance. A bibliography is included. (Author/VM)

ED 056539

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Philip Mansell

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

FL 002 335

Linguistic Parameters in Performance Models

Philip Mansell

1.0 Introduction

Recent work in phonetic modelling has assumed the existence of previous processing; that is to say, it has denied the autonomy of phonetics. In particular it has assumed the existence of a phonology. Under current allophonic theory (Tatham [1969]; see also Kim [1966], Ladefoged [1967]) the *phonology* is to generate those ('extrinsic') allophones which are viewed as part of the linguistic structure of the language, whereas the result of the *phonetic* processing is seen to be the realisation of this language-specific set of extrinsic allophones through a filter, as it were, of 'intrinsic' allophonic rules, which state universal consequences for the intersection of a certain speech intention, time, and the mechanical apparatus of speech.

The general position, then, is clear[1]; what is not clear is:

(i) how the belief that phonetics is not autonomous is derived, and

(ii) the details of the form of the interface between the presupposed phonology and the phonetic processing.

Now, should it be that no good case could be made for the belief that phonetics is dependent on prior processing, then what will be seen below to be the difficult set of problems entailed by the phonology/phonetics interface could be bypassed, and phonetic modelling of a general kind would come into line with the numerous schemes for rule-governed speech synthesis which assume an essentially phonemic input. And, indeed, I think it is true to say that initially allophonic theory lacked a formal proof of the non-autonomy of phonetics. At best, it could point to the fact that, given that the study of performance was in some sense the study of the linguistic form of the utterance in interaction with the environment (a view that will continue to be propounded here) it was the linguistic form (originally identified with the Chomsky-Halle competence model) which was the most richly specified, and hence a good basis for experimental research.

Recently, however, a paradigm has been defined over a number of speech parameters which suggests that we must allow for the possibility of the involvement of linguistic parameters at the heart of a production model for speech (see 2.3.3 below). This paradigm lacks as yet experimental validation, but is susceptible to it. Its very plausibility, however, makes it possible to continue under the overriding assumption that phonetics is not autonomous.

Given this assumption, we cannot bypass input problems; nor are such problems explicitly catered for in allophonic theory, for, as hinted above, allophonic theory can be seen as stating what would be the case, given a competence phonology interfaced with a competence phonetics.

This paper will attempt to demonstrate that it is possible, without an undue number of prior assumptions and without employing circular arguments to proceed towards an independent specification of the input, *via* a model of phonetic processing derived from observation of the output data.

The output from a particular competence phonology having been found wanting in terms of this independent specification of input necessities, the broader question of whether we are in a position to generalise the result of the above arguments and show that the output from *any* competence phonology would fail to meet such input specifications is discussed.

Finally, some experimental consequences of the above discussion are pointed out.

2.0 Specification of Phonetic Processes

There are, admittedly, wide areas of disagreement amongst researchers in the field, but the most striking fact about the three models of phonetic processing currently the subject of experiment (Fromkin [1966]; Cooper [1966]; Tatham [1970b]) is their unanimity with regard to the general tasks of a phonetic component. As stated above, however, it is not at all clear what the theoretical status of these generally agreed notions is. Fromkin [1966] perhaps comes closest to a discussion of such subjects, but there is in her paper something of a jump between highly general considerations relating to model-building, and the specification of a detailed model of phonetic processing.

What is clear is that we are unfortunately far from the conventional 'black box' situation, where the input and output of a particular device are known, and the task of modelling that device consists in selecting from the family of systems that have the same transfer function as the original device. In our case, both the form of the device and the nature of the input are unknowns, while our data on the output are in effect a collection of subsets of data, each subset depending on the experimental techniques used in deriving the data. Certainly we are in possession of more of less complete sets of transformations relating certain of these subsets to others (the prime example being that set of transformations which relates articulation to acoustic output - though even here several potentially crucial topics remain without solution, especially that of compensatory articulation (see Fant [1960]). But other, possibly more vital, sets of transformations are almost entirely

lacking, especially that set that, hopefully, relates electromyographic activity to articulation. And until a complete set of transformations can be derived demonstrating the equivalence between these various sets of data, we have in principle no means of knowing whether our particular characterisation of the output is a suitable one for the purposes of making inferences about the nature of prior processing.

Given such an intractable situation, and given that we have no means of direct experimentation in this area, it is necessary to proceed with a degree of formality.

Initially, then, I want to make certain assumptions about levels of function withinany schema for the production of speech. I shall not produce evidence for these assumptions, but expect them to be non-controversial, and would add that evidence against them - interpreted simply as specifying necessary levels of functioning - would be hard to imagine. The assumptions are:

(i) That there is a phonological level of processing - that is to say, processing which reflects the sound structure of the language, the rules associated with which are not affected by variables associated either with the motor system or with the environment, where the environment can be defined as including the output of, say, affective centres within the brain.

(ii) That there is a level of processing which we can call that of motor planning. I thus separate the central core of phonetic processing from the consequences of such processing, also within the purview of the model, but represented as assumption

(iii) That there is a level of processing which can be seen as executing the motor plan derived under (ii).

(iv) That there is a level of control over the execution of the motor plan.

2.0.2

Of the interrelationships between these assumptions, it is possible to note at this stage: first, that the dependency of (ii) on (i) reflects the overriding assumption stated in the introduction, that phonetic processing was non-autonomous. And second, and, important from the point of view of limiting the discussion below,

(iii) and (iv), the execution of the motor plan and the control over that execution, can effectively be taken as *included* in the level of motor planning for the purposes of model-building, though of course antecendent to it in a physical system. The explanation is as follows: the level of motor planning can be said to include the execution of the plan, because certain of the operations of the motor planning device will undoubtedly require knowledge of the likely results of plans at a lower level; hence the higher device can be said to include at least a partial model of the lower device. To anticipate, this point will be seen as of special relevance in those cases where specific action must be taken by the higher device to counteract an effect which would otherwise inevitably occur as a result of the operation of the lower device. Further, the level of control over execution can be said to be included in the motor planning stage, since we suppose that the form of commands given out from this level will be constrained by the prior knowledge of the control to be exercised over the execution of those commands, given that the degree of control is proportional to the amount of feedback expected.

In effect, therefore, the value of the discussion is not limited by being restricted to levels (i) and (ii).

2.1

I want to consider now what might be inferred about the nature of the processing within the phonetic component, given the above limited set of assumptions, together with certain non-controversial aspects of the input (aspects which are also to be seen as constant over most subsets of the output data). Of the two examples taken, the first is obvious, though not trivial, while the second is more problematical.

2.1.1

We notice first that the output, speech, in whatever form it is sampled, is a continuous function of time. We further notice that, although it is possible that certain phonological features (notably that of *length*) might have as realisations an aspect of measurable time, it is nonetheless to be expected that a phonology meant to capture the generalities of the sound structure of a

language, with a terminal representation neutral with respect to conditions of utterance - which is what the assumption above implies - will never require rule environments specified in terms of measurable time. (This argument is further strengthened by the observation that whereas phonological features of length may be shared in a community, real time realisates of these features will vary widely between individuals, and within the speech of a particular individual generally, and under different conditions in particular. This last point anticipates a later stage in the argument.)

Given that the neuromuscular system is interpretative, we conclude that the transformation of phonological generalities into a form that can be realised in real time is a function of the motor planning device.

2.1.2

The second characteristic of the output I wish to draw attention to is the unit of organisation which is manifest in the output. We are accustomed to label this unit the syllable.

The question to be decided is whether to regard syllabification as a property of the motor planner itself, or as a property of the input from the phonology.

Now, it can be shown that, whatever the genesis of the syllable, syllabic processing is necessary at the motor control level. For if we ask: why should it be necessary for the output to manifest syllabic form? the traditional answer might be: to relieve memory load. And if we further ask why memory load should need to be lifted at precisely this point, the answer is surely that there are motor plans which need to be made with respect to a considerable stretch of the utterance - these include those plans relating to intonational phenomena, speech rate and so on. Thus, for the purpose of such operations, one can envisage a situation for a segmental input where the incoming segments are fed into a series of syllable forming rules and are then buffered such that the suprasegmental timing rules can operate.

It can further be observed from the output that the ostensible syllable structure of the utterance appears to be changed with increasing speech rate. I have not yet derived a means whereby phonetic processing can be made

dependent upon external criteria such as speech rate (see below) but even at this stage we can hypothesise that syllabification and time-conversion are processes that are very closely allied. In fact, in any general account of the operation of external variables upon speech parameters it will obviously be of interest to consider the influence of the particular variable of speech rate as bearing directly upon the syllabification rules - and this is in effect what is suggested in Tatham [forthcoming].

These arguments hold under the assumption of a segmental input; but for obvious reasons I am not prepared to admit assumptions about the form of the input into this argument. Reverting to the original question posed in this section, then, on what grounds might we suppose the syllable to be a phonological entity, and how would such a supposition affect our view of the internal workings of a phonetic component?

The phonological status of the syllable has been the subject of debate (Kohler [1966]; Anderson [1969]; Fudge [1969]); it has been noted that the syllable appears to be an implicit notion in segmental phonologies such as that of Chomsky and Halle [1968]; and in their Akan phonology Schachter and Fromkin [1968] make explicit use of a phonological syllable.

Obviously even a partial consideration of the status of the syllable within phonology would be far beyond the scope of this paper, but the point is crucial for the reason given above, namely that I do not wish to make assumptions about the nature of the input. A convenient solution to this dilemma would be to show that whichever of the two alternatives - syllabic or segmental input - were chosen the effect on the form of the motor planning device would be minimal. If this could be shown, then it could further be shown that the type of input problems at issue later in this paper remained problems whether the input is specified asbeing segmental or syllabic in form.

2.2

In order to approach such a demonstration it is necessary first to make a general observation on our expectations of *any* input derived from *any* preceeding phonology; and that is that we expect even under the most ideal interface conditions that the input will be

in some measure inappropriate for the motor processing. Here one would define 'inappropriate' as 'lacking information that we can properly demand that the phonetics supply'.

To take the most trivial set of examples: that of the labelling of features. It would seem that we must build into phonetic processing a series of readjustment rules which map phonological features into parameters of muscular involvement. (Note that this will be the case even if the phonological features are expressed in direct articulatory terms.) Now there is no reason to suppose that the mapping will be one-to-one, or that the input may not be redundant in that involvement of the same muscle may be signalled twice in the couse of a single segment. Further possibilitie exist: the presence of a particular muscle may be signalled by the simultaneous presence of two or more features. A single feature may lack realisation; or may act only as a conditioning environment for the realisation of another feature by a particular muscle. The same possibilities exist, of course, for the specification of the absence of involvement of a particular muscle in a given articulation[2].

In the motor planning processing envisaged for the case of a segmental input, then, we have a series of readjustment rules, followed by syllabification rules, after the operation of which the string is buffered and operated upon by suprasegmental rules. What would be the effect on this configuration of a syllabically organised input? Given such a minimally specified set of processes, of course, it is not possible to predict the effect in great detail. But the following points can be made:

(i) I presume that even a syllabically organised input will still represent segments within syllables. Hence the readjustment rules will still necessarily apply.

(ii) It is unlikely that the motor specifications of the segments will be altered by their syllabic environments at the stage of the readjustment rules, for these environments will be phonological in nature, language-specific in reference, and will have been used, I infer from the phonology of Schachter and Fromkin, to state sequential constraints on the combination of segments[3]. The syllable at the motor level, on the other hand, the organisation of which may well affect the individual motor composition of the segments entering into combination, are motor entities and language universal in nature.

It may thus be argued that acceptance of a syllabic input has the single consequence that we must change the set of syllabification rules into a set of syllable-deformation rules. I would suggest that this represents a minor modification, and that it is thus possible to regard the syllable/segment dichotomy as of secondary importance in the general framework of input problems.

2.3

The model outlined in the previous section is not complete even in very general terms; nor have we reached the end of what we may legitimately infer about the nature of phonetic processing on the basis of the initial assumptions and observations on the output. Even if the minimal model previously described were supplied with detailed rule frameworks to carry out the functions postulated, its output would fail to model our data on real speech.

The principal features of speech output not represented in the output of the minimal model are all related to the variability of the output, and are as follows:

2.3.1

We observe that speech is variable in a manner which suggests that a number of the processes it has undergone are governed by factors external to the linguistic situation - which factors I will call *environmental* factors, where environmental, as above (section 2.0.1) is taken as including 'emotional'. I have already noted the apparent change in syllable structure with speech rate (see also Kozhevnikov and Chistovich [.965]). One could also point to the rise in overall speech intensity as a function of external noise level (Lane *et al.* [1970]) and to changes in speech rate as a function, say, of urgency or excitement. Perhaps a paradigm example of reaction to environmental factors is the phenomenon of whispering, which is discussed in greater detail later.

It is to be noted that it is rarely the case that the variation brought about by environmental factors is either simple in form or limited in extent; change in speech rate, for example, does not result in simple compression or expansion.

2.3.2

Secondly, speech production is seen to be variable between successive utterances of what is linguistically the same stimulus item, in cases where environmental factors cannot reasonably be appealed to. This variation may be of more limited extent than the first type.

2.3.3

Thirdly, such a minimal model would predict identical output for the centre element of phonologically identical VCV or VVV utterances in two different languages, irrespective of the overall structure of that language. It will be remembered that under the terms of allophonic theory as sketched above the consequences of this would be the prediction that the same degree of mechanical co-articualtion would occur between these centre elements and their environments in both languages, these mechanical co-articulations being known as intrinsic allophones.

We return here in fact to the paradigm example in allophonic theory which I mentioned earlier, on which rests the possibility of a formal proof of the non-autonomy of phonetics. For allophonic theory now predicts that the degree to which natural mechanical co-articulation is allowed to affect a given segment in a given environment is an indirect function of the structure of the language in question. Thus Morton and Tatham [1970] claimed that in a language with only four vowels with which to differentiate morphemes, say /i/, /æ/, /ɑ/ and /u/, the degree of co-articulation for the [i] in the sequence [ɑiɑ] would be greater and would more nearly approach the natural degree of co-articulation than in a language with twelve vowels, including in particular /i/, /ɪ/ and /e/. Again, for a CVC utterance in which the C is a palatal fricative, one of two in Language 1, and the two V's are back vowels, Tatham [1970b] predicts greater co-articulation than in Language 2 which possesses segments identical to those of Language 1 but where the palatal fricative in question was one of four such fricatives in the language.

2.4

Now, of these three features of the speech output, the first, the dependence of speech parameters on aspects of the environment, will be seen to imply a radical reorganisation of the internal organisation of the phonetic processing mechanism. When this reorganisation has been effected, then the two other features of the output, the one well attested, the other only hypothesised at present, call, for their adequate modelling, for information which we cannot reasonably suppose it the task of the phonetic component alone to bring to the task of synthesising utterances. Consideration of these features of the output, then, will bring us back to the question of the desired nature of the input to a phonetic component. And, in fact, I shall be discussing observations (ii) and (iii) in detail only at this later stage.

2.4.1

In the original assumptions of this paper, the motor planning stage of the phonetic component was seen as a function only of the phonological input; the resultant model (our 'minimal model' of the previous section) could thus be characterised as a *linguistically-oriented model*. As result of observations on the speech output we would now wish to represent the motor planning as a function *both* of the phonological input and of the environment, this being the formal analogue of the theoretical characterisation of performance as the intersection of the linguistic intention and the environment.

Suppose for the moment that the influence of the environment is conceptualised in quite a general way as a single control element to be introduced at a point in the processing already hypothesised for the 'minimal model'. Now, the point requires arguing at length, but, in terms of our control element conceptualisation, I can see no formal reason for introducing the influence of the environment at any particular point in the hypothesised processing, other than either at the beginning or at the end. That is to say, at any given point there will always be either preceeding or subsequent processes which will both require explicitly stated controlling conditions, and I can see no reason why we should initially choose to construct an element with a mixed form of rules. There is a stronger form of this argument, and one which affects the plausibility of the suggestion that the control element

should be placed at the end of the hypothesised processing; and that is that we should be lead, on the grounds of computational efficiency alone, to reject any rule which involves backward control, or which, in other words, involves the resynthesis in terms of an external criterion of what has already been synthesised, presumably in the absence of such a control.

However, a criterion of efficiency of computation could take many forms, and is therefore not perhaps a sound criterion on which to base this argument. The position would be immeasurably strengthened if it were possible to show that, once the principle of phonetic processing being governed by external criteria had been admitted, it is impossible to model a system of phonetic processing adequately without having all the products of such a system modified by such constraints. Such a demonstration would show the fallacy of backward control by denying the possibility of any computation free from external constraints. I shall undertake the demonstration with the aid of an example.

2.4.2

The example is that of whispered speech. The example is apt, in that the occurrence of whisper can be seen as directly related to environmental pressures on the speakers. But at the same time one is hampered by a paucity of data on this subject. As will be well known Meyer-Eppler [1957] hypothesised a number of changes in the spectral composition of vowels to compensate for the absence of a means of signalling pitch changes in whispered speech. The conclusion that this is a regular, conventional procedure has recently been questioned by Fonagy [1969]. Kloster-Jensen [1958] examined the perception of lexical pitch variations using examples from a number of languages. The laryngeal adjustment for whisper is well known (see e.g. van den Berg [1968]). On the phonological level Fries and Pike [1949] have some discussion on possible means of handling whisper within a phonemic system.

These studies, however, have dealt with only restricted aspects of the articulation of whispered speech, and even here no clear picture has emerged. I shall limit my discussion of whispering in the following ways:

(i) I shall assume that the phonological representation of an utterance is invariant, whether this utterance is realised in voiced or whispered mode. Evidence for this claim is given by Sharf [1964] who shows that those aspects of vowel duration which are dependent upon the phonological specification of [±voice] in following consonants are maintained in whispered speech.

(ii) I shall restrict the problem to a consideration of segments marked phonologically as [+voice] (lacking an information about the articulatory configuration of the larynx during segments marked [-voice] in the phonology)[4].

Further, as well as observing that whispering appears to be a response to environmental pressure, I shall assume that it has been observed that the occurrence of voiced sequences is very much more frequent than the occurrence of whispered sequences; and hence that phonetically we would wish to designate whispered speech as the marked occurrence, voiced speech as the unmarked occurrence.

2.4.2.1

Now, under the readjustment rules within the phonetic component that were introduced above, let us assume that the phonological feature [+voice] is transformed into a feature specifying, say, *adduction of the vocal folds*. We may then represent the choice between voicing and whispering as between two modes of adduction of the vocal folds, say, *vibratory* and *uncoupled*. Now, if we were setting out a model, contrary to hypothesis, where voiced sequences could be realised without being subject to external constraint, then it seems necessary to postulate that the state symbolised as *uncoupled* would need to be derived from the state *vibratory* by an optional transformation, with the environment specified only in this transformation. Such a state of affairs clearly involves a confusion of criteria: for although it is observed that voiced sequences are of more frequent occurrence, I know of no independent argument which could show that the two possibilities are to have different rank orderings within the model, which would be the result if the one were derived from the other. Clearly the states *vibratory* and *uncoupled* are both realisations of the feature relating to adduction of the vocal folds, and are to be seen as equivalent in terms of the rank ordering of the model;

what are not equivalent are the probabilities of occurrence of the presence or absence of the environmental constraint relating to whispered speech.

I would contend that the only system in which we can satisfactorily model *both* the rank equivalence *and* the distinction between marked and unmarked forms with respect to the two postulated realisations of the feature of adduction of the vocal folds, is to model the unmarked term as also being subject to an environmental constraint, the constraint that the condition which would lead to an utterance being whispered is not present[5].

In order to implement this conclusion by means of rule arrays, we would need to envisage If-Then rules which referred to the existence or non-existence of environmental constraints. In order to provide a formal definition of the marked/unmarked distinction, we should require that a device first access those rules referring to the absence of a condition.

2.4.3

To sum up, then, I hope to have shown by means of this example the fallacy of backward control, and hence to have justified the rejection of the insertion of environmental constraints, viewed as a single control element, either at the end or in the middle of the linguistically-oriented processing of the 'minimal model'.

It has not been proved, of course, that the control element should be placed prior to this linguistically-oriented processing; and in fact this proves not to be necessary, for having used the single control element conceptualisation in order to show the fallacy of backward control, it is now possible to reconsider the notion, and in doing so effect a radical change in the shape of the phonetic processing, rather than a simple addition as would be the case if we adopted the single control element conceptualisation. For consider what processes would be necessary if there was indeed a single control element before the linguistic processing. We would need to bear in mind that specific environmental constraints might well refer to highly specific parts of the subsequent linguistically-oriented processing. Hence we should need to postulate a mechanism such as that adopted by Chomsky and Halle [1968] for the treatment of exceptions: the attachment to segments of further features specifying

(in our case) the environmental feature which will govern the form of processing at some point later in the model. Such a mechanism would need to be combined with sets of alternative rules, labelled in terms of the features they state the consequences of (or having that feature specified in the If part of an If-Then rule) stored at the processing locations later in the model.

The system for rule specification is only what we have predicted before, but the mechanism of attaching features to segments in order to specify the rules they are subject to later in the model seems an entirely unmotivated artifice, which has the simple effect of increasing the computation load, in that more complex matrices are produced, some containing features which will perhaps need to be carried unnecessarily through a number of processes before being utilised to effect a selection from a choice of rules. It may further be noted that such a system would lead to unnecessary reduplication of feature specification in the case of environmental features relating to prosodic features of the utterance.

The solution to this dilemma lies, I believe, in abandoning the form of the minimal model and its orientation to the linguistic form of the utterance and viewing the entire processing, as a whole and in its separate parts as a function of environmental features. We should thus allow the environmental features to enter the model at the point in the processing where they are relevant, keeping the notion of rule arrays specified in terms of the presence or absence of such features. Viewed as a single device, the phonetic motor planner will thus have a number of inputs, of which the phonological form of the utterance will be only one. The output from the device will reflect the state of activity at all these points.

2.4.4

This, then, is the final form of the model of phonetic processing in terms of which I wish to discuss the optimum form of input. The model as it stands makes certain demands with respect to its input, but these can be more clearly demonstrated in terms of the two apparently contradictory observations about the extent of variation in the output which were introduced earlier in this section.

3.0

The two observations regarding variation in the speech output not already dealt with were:

(i) That for two different languages with identically specified segments at a phonological level, in identical phonological environments, the degree of mechanical co-articulation for those segments would differ according to the total sound pattern of those languages.

(ii) That for a single speaker on a given occasiona, repetitions of a phonologically identical utterance will not have the same physical realisations.

(i), it will be remembered, has the status of a hypothesis, while (ii) is well attested in a number of subsets of output data (see Mansell [1970] and the references there). It is interesting, therefore, that both types of variation can be handled by the same mechanism.

3.1.0

The hypothesis concerning differing degrees of co-articulation requires that information related to the structure of the language be available at the level of phonetic processing. The information contained in a single column of a distinctive feature matrix could be said to contain information not only about its own identity, but also, by inference from the labels of the rows of the matrix and the binary or integer notation in the cells of the matrix, information about all other segments. These 'other segments' would however be all other segments describable in terms of that particular matrix, irrespective of the pattern of that particular language. And even could we first derive inferential rules and second constrain such rules to produce the desired information about a particular language, it is doubtful whether such information would be of great practical importance in modelling this aspect of the output in terms of a model such as that described in the previous section.

For in such an environment-dominated model, the linguistic information that is to modify the motor pattern must be expressed in terms of the *communicability* of the sound pattern of a language. In short, we require from the *phonology* information relating to the relative importance of segments and parts of segments in the sound

structure of the language, information which is to be used in the *phonetics* to provide answers to such questions as: given this communicative situation (i.e. the totality of environmental pressures) what are the features which will enable me to distinguish this segment from others (not necessarily all others) in the language. It is interesting that Morton and Tatham (*op.cit.*) and Tatham (*op.cit.*) introduce the psychological principle of maximum differentiation to explain the overriding of natural co-articulatory tendencies that they hypothesise. What a phonetic model as envisaged here must demand from its input is a formal analogue of just such a principle.

This requirement in effect is a collapse of two separate requirements: the first is that the labelling of the rows of the matrix, to put the problem in purely graphic terms, relate to the communication of the structure of the language. About this requirement I have nothing substantive to say; the details of the re-adjustment rules postulated within the phonetic component are as yet so little worked out it is impossible to say whether an intermediate level between classificatory and motor features (which is what in effect features relating to communication would be) is in fact necessary. The second requirement relates to the need for knowledge about the motor output relevant to a particular environmental load: in other words it requires hierarchical structuring within the columns of input matrices, the hierarchy to provide a rank ordering of features relative to their necessity of inclusion in the output in order that a particular distinction be maintained.

3.1.1

Turning now to the other instance of variation, that between successive utterances of phonologically identical items, experimental evidence appears to show that this variation takes place on only a limited scale (though this may be a false impression induced by the fact that variation has not generally been investigated for its own sake, but has been reported on as a result of specialist investigations). Thus, on the single parameter of lip protrusion for the vowel [ʊ] Mansell and Allen [1970] reported preliminary evidence that successive articulations differed quite widely, although the phonological context was constant, and the subject was instructed to keep his utterances as similar as possible.

It is interesting to note that the subject was able, as a result of this instruction, to keep both duration and amplitude constant; and it is further of note that spectrographic evidence did not show discernible spectral variations with the variations in lip protrusion.

It would seem that, in order to explain data of this nature, we are again led to the concept of hierarchically ordered features within segments. For, just as it was argued that such ordering could provide an index of what was necessary for communication, so it can be argued that it can provide an index of what is *not* necessary. In particular it was proposed (Mansell [1970]) that some features of segments could be viewed as being at such a low level in the communicative hierarchy that their value within a particular language pattern could be said to be almost negligible. Control over their production would therefore be at a minimum, and the value reached on some particular measurement parameter applied to the output would be in almost perpetual oscillation.

3.2

It is evident that the requirements upon an input stated above are not met by the output from a Chomsky/ Halle phonology. It will be noticed that the inadequacy has not been demonstrated in terms of the basic form of the input - segmental or syllabic - nor even in terms of its matrix representation - viewing the task of phonetic processing as eliminating the rows and columns of a matrix can be a formal analogue of the dual tasks of converting linguistic information to motor information and adding measurable time to the representation. Rather the objection has been to the terms of reference of such a putative input; and it has been suggested that the reference should principally be to the communicative function of the language structure, such a reference being made principally by means of hierarchical ordering within segments.

What is required in fact is a performance phonology organised in such terms; but lacking such a phonology it will be necessary to continue the process of inference from speech output data until we have a more viable and complete specification of the input data required. I hope in this section and the last to have shown that such an undertaking is possible, and, in that it leads to generalisations over a number of aspects of the output data, desirable.

4.0

The question could be left at this point; but mention has been made of a performance phonology, and I believe that it is possible to add to the discussion by considering aspects of the competence/performance dichotomy. In particular, I wish to see how far one can generalise the above result in order to show that no output from any competence phonology (and since we are using competence as a technical term this must mean any variety of TG phonology) would satisfy the input requirements of a phonetic component.

4.1

With regard to the first question, it does not appear possible at this stage to generalise from the particular conclusion given in Section 3. A number of points may be made, however, which, it is hoped, will show those areas where we lack the knowledge that alone would make a decision on this question possible. It is to be noted that the conclusions that Watt [1970] has recently reached in this area rest on the basis of a direct comparison between experimental evidence and predictions of complexity derived from a competence model. These conclusions were, briefly, that a competence grammar, as defined by the theory, could not in any of its conceivable forms, be either included in or correlated with the mental grammar that a hearer employs to decide syntactic material. But that the sub-optimal grammar of the adult was nonetheless a grammar.

Perhaps the best means of approaching this question is to consider a performative schema for the production of speech which will include processing in those areas we conventionally designate phonology and phonetics. It is possible then to ask, like Watt, whether the assumption that a competence model (in this case the phonology) is to be included in a performance model (the schema) is valid. Chomsky's statement of this assumption is well known and reads as follows:

> No doubt, a reasonable model of language use will incorporate as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the language; but the generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production.
>
> (Aspects, p.9)

Now, it is evident that this basic assumption could be made a good deal weaker if it were possible to show that a competence model *does*, in itself, constrain the character and functioning of a performance model. It seems to me that the rudiments of such a demonstration do exist, although lack of information does not enable us to carry the argument through. Here we need to make use of the descriptive term frequently applied to competence models - that they are *monolithic* in form. This means in practice that the information stated in a competence model can only be accessed *en masse* (and this effectively means at the start of a performance model) or in blocks according to the natural breaks in the model. Now, it has already been seen that accepting the competence model as input to the performance gives rise to problems as to the specification of this input. For a convincing argument it would need to be shown that such input problems were ineradicable given the form of the competence model, *and* that the alternative, dispersion in blocks, was at the least unmotivated.

Hence a proof of the generic unsuitability of competence models within performance would have to show:

(i) That the natural breaks in such models are of no behavioural significance,

(ii) That the form of the output at present employed is necessary to the operation of the present and any future competence phonology,

(iii) That the introduction of such information as a phonetic component might demand into a competence phonology would not be justifiable in terms of the general constraints that such competence models are subject to. (This last being closest to the types of arguments used by Watt [*op cit.*].)

4.1.1

It is obvious, I think, that we are in no position to demonstrate either (i) or (ii).

The only theorist to my knowledge to have considered the behavioural relevance of the breaks in the phonological model is Halle [1962] (though Kiparsky [1968a, 1968b] has made claims for the behavioural relevance of other aspects of phonological formalism), who uses the divisions of a competence phonology in the process of discussing an 'intelligibility criterion' for the introduction of innovations into the sound patterns of languages. He writes:

> I am unable at present to characterize the place in the order where rules may be added with a minimum impairment of intelligibility. Such additions, however, seem characteristically to occur at points where there are natural breaks in the grammar.

I know of no claims to the contrary, or of further claims denying the behavioural relevance of breaks in the grammar.

Further, (ii) seems implausible. Wilson [1966] has specific arguments that aspects of generative phonology are not functions of the adoption of distinctive feature matrices, as Halle [1962] appears at one point to claim, but have a quite separate basis for validity in general linguistic theory. And in principle, I can see no reason why a phonology should be constrained to a particular phonetic representation, except in terms of (iii) below. It is interesting to note in this context that Chomsky [1964, p.120] proposed the adoption of a universal phonetic alphabet for the very utilitarian purpose of satisfying the requirement that a grammar should have a set of output sentences.

(iii) would appear, then, to be the most profitable line of investigation here; and it indeed seems possible that a competence phonology could not be found that could justify by internal criteria the representation of such information as the phonetic processing appears to require. I am not at present prepared, however, to use the relatively meagre set of input criteria derived for phonetics for this purpose.

4.2

It is unsatisfactory to conclude this discussion on such a negative note, however. It is the case that we need to model phonetic behaviour on a level beyond that of individual behavioural acts; what is becoming clear, in phonetics as in psycholinguistics (see Watt, *op cit.*; Fodor and Garrett [1966]; Vandamme [1968]) is that competence models are particular constructs within linguistic theory which are not necessarily related to the task of modelling performance. Watt [*op cit.*] has derived an Abstract Performative Grammar on the basis of his experimental evidence. It is to be hoped that further attempts to specify the nature of the input to the phonetic component, and to systematise the form of the rules within that component, may eventually lead to a similar solution in phonetic modelling.

5.0

Very briefly, those areas of the above discussion which require experimental evidence with some urgency are those of *alternation*, *variation* and *co-articulation*.

5.1

The treatment of alternation in the model has been investigated by means of the example of whispered speech. It is imperative in order to add substance to this treatment, that an investigation of articulatory activity during whispered speech be carried out. It is suggested that only with a knowledge of articulatory parameters can we hope to approach the question of compensatory variations in the acoustic output with any degree of success.

5.2

The study of the variations between successive utterances of phonologically identical tokens is already in progress; it was stimulated initially by an investigation of various unsatisfactory features of electromyographic technique in speech research (see Mansell [1970] and Mansell and Allen [1970] for details). This research is continuing and has received an additional impetus from the theoretical interest of the topic.

5.3

A large part of the argument above rests on a hypothesis concerning extent of co-articulation in different languages. Experimental validation of this hypothesis must wait on the setting up of adequate parameters to measure co-articulation. The work of Ohman on the tongue has perhaps constrained us too heavily into thinking of co-articulation as being a predominantly two-dimentional phenomenon.

6.0

Conclusion

This paper has dealt primarily with problems concerning the nature of the input to a phonetic processor; various criteria for such an input relating to the *communication* of the sound structure of a language were derived as a result of a specification of that phonetic processor on the basis of a limited number of assumptions linked to inferences from speech data.

It was further suggested that such procedures were of theoretical as well as practical relevance in that they make it possible to raise larger questions concerning the interrelation of competence and performance.

Notes

[1]Clear, that is, if either one of two simplifying assumptions is made:

(i) That the set of 'extrinsic allophonic rules' is coextensive with the set of phonological rules, given in some phonology.

(ii) That the notion 'the set of extrinsic allophones' is defined simply as coextensive with the given set of outputs from a presupposed but not necessarily characterised phonology.

Assumption (ii) may be held, of course whatever content we might wish to give to the notion 'extrinsic allophone' (and it will be seen from the text that the view preferred here does not correspond to the output of any known phonology); though this assumption may be a convenient one for this reason, it is evident that any characterisation of the notion 'extrinsic allophone' would gain in interest if constrained from two directions, by the nature of both phonological *and* phonetic processing.

The adoption of Assumption (i) may be justified by the emergence of some future phonology; for the present the testing of this assumption against an existent phonology is interesting, not necessarily for the nature of the result, but for the purpose of deriving the grounds on which we might wish to partition a set of phonological rules. For one such attempt and attendant reasoning, see Postal [1968] (Universal Detail Rules, p.66 f.).

These questions are not explicitly discussed in the body of the text, but would be of great importance in a wider treatment of the phonology/phonetics interface.

[2]I am here regarding phonological matrices as capable of being transformed only into matrices where obligatory presence, or absence, or neutrality is shown for particular muscles or muscle groups. Given that all information relevant to motor planning is extracted from the input string, I would wish the specification of *amount* of muscle contraction in a given instance to be a subsequent, purely motor, decision. In this view, the phonological specification by means of integers of degrees of features (with the exception, of course, of stress) would only be necessary in order to change which muscles were selected.

[3]This argument does not hold for the Lugisu phonology given by Brown [1969]. She writes (p.6):

> The syllable we are concerned with here is not ... an underlying distributional unit, a phonological syllable, but the distributional unit of the phonetic realization.

It is beyond the scope of this paper, however, to discuss the details of Brown's proposals.

[4]The question to be decided experimentally is whether we can represent the whispered mode of speech as differing from the programme for voiced mode in only extremely limited ways, or whether we must hypothesise a communicative strategy for whisper which has such global effect that we need to consider it to be an entirely separate programme on the phonetic level.

This issue is basically that which is examined with respect to the communication of pitch information in whisper in Meyer-Eppler [1957] and Fonagy [1969] (see text). The same problem, however, presents itself in a much simpler form when we come to consider the sequence of events at the larynx during the utterance, say [ɑpʰɑ] in voiced and whispered modes. If we were to hypothesise strictly local deformation of the total programme for this utterance in the case of whispered speech, then we should expect redundant abduction of the vocal folds during the closure for the phonologically voiceless stop. If, on the other hand, we hypothesised a different global strategy, we should expect, not only that *ad hoc* manoeuvres would be undertaken to simulate the communication of features only directly communicable in voiced mode (pitch information again), but we should also expect that manoeuvres necessary in voiced mode but redundant for whispered speech would be avoided. Hence we should expect to find no abduction manoeuvre of the glottis during voiceless stops.

It is interesting to note here from the point of view of the general form of phonetic explanations that there are two criteria of simplicity involved in these conflicting hypotheses, which criteria are themselves in conflict. That there should be only limited differences in the case of whisper, and hence redundancy evident in the articulations, and reliance on secondary cues, meets a criterion of simplicity relating to the overall form of the phonetic component. That there should be a separate strategy, involving the excision of redundancy and the employment of extra whisper-specific manoeuvres,

would fulfil a criterion of simplicity just for the limited case of whispered speech.

It will be seen then that experimentation concerned with glottal manoeuvres during speech is of quite general relevance. Such experimentation is at present under preparation in this laboratory.

[5]The formulation requires further explication for the following reasons:

(i)The notion 'rank ordering' is imprecise.

(ii) The rejected device of deriving the marked form from the unmarked by means of an optional transformation is an approach conventionally used in the syntactic component of a generative grammar.

In the case where Z is the marked realisation of Y, and A the unmarked realisation, suppose the alternation rules:

1. If $\underline{X}$, then $Y \rightarrow Z$
2. If $\overline{X}$, then $Y \rightarrow A$

Suppose further that these rules have derivational consequences:

3. If Z, then $(Z_1, Z_2, Z_3, \ldots, Z_n)$
4. If A, then $(A_1, A_2, A_3, \ldots, A_m)$

where Z_1, Z_2 etc. and A_1, A_2 etc. are steps in the derivation of Z and A respectively, and Z_n and A_m belong to the terminal alphabet of the derivation, and $m<n$.

Then the differences in complexity between rules {1,3} and rules {2,4} lie either in the *If* parts of {1,2} or in the *Then* parts of {3,4} or both. I have shown the case where the marked realisation is derivationally more complex; this would be conventionally held to be a justification for the choice of Z as the marked term. But if we were to graph the comparative derivational history of the two forms, say by entering on a tree diagram the results of successive operations, it would be obviously seen that the non-terminal categories A and Z would have precisely the same degree of derivational complexity intervening between them and the superordinate category Y - in this case they would both be immediately dominated by Y. It is claimed that this equivalence of rank ordering of the alternate realisations of a superordinate category is in principle distinct from the non-equivalence of the derivational complexity intervening between A and Z and A_m and Z_n.

(A brief note on the rules above: For the special case of (2) where A=Y, and hence where for (4) A_m=0, however, there remains a formal problem. Either (2) is to be regarded as redundant, or is to be allowed to apply vacuously. In the first interpretation, the marked form is derived from the unmarked form by default. For opposing views on this point, see Lamb [1964] and Postal [1968].)

Turning now to the case of syntax, it is necessary to examine the status of optional transformations. We can distinguish between a rule system for the enmeration of sentences (a competence grammar), and a pass through those rules (a schematic, though scarcely adequate, account of performance). In the latter case, the existence of an optional transformation constitutes a hypothesis about language behaviour - namely that the variation is random in nature. If random behaviour does not appear appropriate, and conditions are attached to the alternation, then the optional nature of the process is destroyed.

In the case of the enumerative system, I would claim that the existence of optional transformations shows the limit of applicability of the categories defined in the system. Evidently there is a form of trading relationship between the number of categories defined in the system (and hence the number available for the statement of rule applicability contexts) and the number of situations which can be defined within that system. (The total number of situations may be less than the number which could be defined given the number of categories, according to the operation of the evaluation measures chosen, of course; indeed, we may expect to find optional transformations in a system where the number of categories is optimal with respect to a particular evaluation measure for the system as a whole. What is in question here is not the satisfactory or unsatisfactory nature of the devices of particular competence systems, but whether these devices have relevance for the formalisms of a phonetic component.) To take only one example: it appears to be the case that were syntactic theory to be extended to include categories relating to discourse phenomena (Topic, Comment, Focus, Presupposition, etc.) then the optional status of certain transformations (notably the Pseudo-cleft; see Kay [1967] p.9) would be no longer defensible in certain situations defined by these new categories.

In phonetics one is hoping to specify a performance model of much greater complexity than the 'pass through the enumerative rules' referred to above; even in the present initial stages, it seems unlikely that random

variation will play much part in phonetic explanations. Further, the categories of phonetics are at present far from being defined or even listed. The relevance, therefore, of a schema for handling alternation which appears to be a function of a particular limited set of categories is disputed.

Copyright 1970 P. Mansell
University of Essex
Language Centre
Occasional Papers #8

Bibliography

Anderson, J. (1969) Syllabic or non-syllabic phonology? Journal of Linguistics 5, 136-142

Berg, Jw. van den (1968) Mechanism of the larynx and the laryngeal vibrations in Manual of Phonetics ed. Malmberg, North Holland, pp.278-308

Brown, G. (1970) Syllables and redundancy rules in generative phonology Journal of Linguistics 6, 1-18

Chomsky, N. (1964) On the notion 'Rule of Grammar' in The Structure of Language, ed. Fodor & Katz, Prentice Hall, pp.121-136

——— (1965) Aspects of the Theory of Syntax MIT

——— and Halle, M. (1968) Sound Pattern of English Harper & Row

Cooper, F.S. (1966) Describing the Speech Process in Motor Command Terms Haskins Lab, Status Report

Fant, G. (1960) Acoustic Theory of Speech Production Mouton

Fodor, J. & Garrett, M. (1966) Some reflections on competence and performance, in Psycholinguistics Papers, ed. Lyons & Wales, EUP, pp.135-154

Fonagy, I. (1969) Accent et intonation dans la parole chuchotée Phonetica 20, 177-192

Fries, C.C. & Pike, K.L. (1949) Coexistent phonemic systems Language 25, 29-50

Fromkin, V. (1966) Speculations on Performance Models Journal of Linguistics 2, 47-68

Fudge, E.C. (1969) Syllables Journal of Linguistics 5, 253-286

Halle, M. (1964) Phonology in a Generative Grammar, in Structure of Language, pp.334-354

Kay, M. (1967) From Semantics to Syntax, Rand Corporation P-3746

Kim, C-W. (1966) The Linguistic Specification of Speech Working Papers in Phonetics 5, UCLA

Kiparsky, P. (1968a) Linguistic Universals and Language Change, in Universals in Linguistic Theory ed. Bach & Harms, Holt, Rhinehart & Winston, pp.170-202

——— (1968b) How Abstract is Phonology? Indiana University Linguistics Club

Kloster-Jensen, M. (1958) Recognition of Word Tones in Whispered Speech Word 14, 187-196

Kohler, K.J. (1966) Is the syllable a phonological universal? Journal of Linguistics 2, 207-208

Kozhevnikov, V.A. and Chistovich, L.A. (1965) Speech: Articulation and Perception Trans. US. Dept. Commerce

Ladefoged, P. (1967) Linguistic Phonetics Working Papers in Phonetics 7

Lamb, S. (1964) On Alternation, Transformation, Realization and Stratification, in Rep. 15th. Ann. Round Table Meeting on Linguistics and Language Studies, ed. Stuart, Georgetown UP., pp.105-122

Lane, H. et al.(1967) Regulation of Voice Communication by Sensory Dynamics JASA 47, 618-624

Mansell, P. (1970) On the nature of EMG variations, to appear in Proc. Essex Symposium on Models of Speech Production, September 1970 , University of Essex Language Centre, Occasional Papers 9

——— and Allen, R. (1970) A first report on the development of a capacitance transducer for the measurement of lip excursion, to appear in University of Essex Language Centre Occasional Papers 9 (see previous ref.)

Meyer-Eppler, W. (1957) Realization of prosodic features in whispered speech JASA 29, 104-106

Morton, K. and Tatham, M.A.A. (1970) The Phonetic Component paper read to the spring 1970 meeting of the Linguistics Association of Great Britain, in University of Essex, Occasional Papers 8

Postal, P.M. (1968) Aspects of Phonological Theory Harper & Row

Ohman, S. (1967) Numerical model of coarticulation JASA 41, 310-320

Schachter, P. and Fromkin, V. (1968) A Phonology of Akan Working Papers in Phonetics 9 UCLA

Tatham, M.A.A. (1969) Classifying Allophones University of Essex Occasional Papers 3

——— (1970a) Coarticulation and Phonetic Competence paper delivered to the April 1970 meeting of the Acoustical Society of America

——— (1970b) A Speech Production Model for Synthesis-by-Rule Working Papers in Linguistics 6, Computer and Inf. Science Res. Center, Ohio State University

——— (forthcoming) Model-building in Phonetic Theory Language Sciences, Indiana University

Vandamme, F.J. (1968) On problems implied by the concept performance MIT QPR 85, 301-303

Watt, W.C. (1970) On two hypotheses concerning psycho-linguistics, in Cognition and the Development of Language, ed. Hayes, J.R., Wiley, pp.137-220

Wilson, R.D. (1966) A criticism of distinctive features Journal of Linguistics 2, 195-206